***YouTube Data Analysis***



### *ABSTRACT*

This project investigates engagement metrics on YouTube to understand what drives video popularity, particularly focusing on views, likes, and comments. Data was collected using the YouTube Data API, offering insights into trending videos and audience interaction patterns. Descriptive analysis and correlation measures reveal the influence of video characteristics on engagement levels. Findings from this analysis could help content creators optimize their content strategy, enhance engagement, and strategically select the best timing for uploads.

### *TABLE OF CONTENTS*

1. Introduction - pg 2
2. Data Description - pg 2
3. Data Preprocessing - pg 3
4. Exploratory Data Analysis (EDA) - pg 3 -4
5. Results and Discussion - pg 5-7
6. Conclusion and Future Work - pg 7
7. References - pg 8

### INTRODUCTION YouTube has become one of the most popular video-sharing platforms worldwide, drawing billions of users and serving as a significant tool for content creators, marketers, and researchers. Analyzing engagement metrics—such as views, likes, and comments—provides valuable insight into user behavior and video performance on the platform.

**Objective:**This analysis aims to explore YouTube engagement metrics using data collected through YouTube's API. The goal is to understand the relationship between these metrics and determine key factors influencing video popularity, ultimately offering practical insights for optimizing content strategy.

### *DATA DESCRIPTION*

**Data Source:**Data was obtained using the YouTube Data API, accessed through [Google Cloud Console](https://console.cloud.google.com/welcome?project=precise-data-392110&pli=1), covering key metrics such as view count, like count, comment count, and other video-specific details.

**Dataset Overview:**The dataset includes recent video trends and metrics, with attributes capturing viewership and interaction patterns.

**Variable Descriptions:**

* view\_count: Integer, total number of views.
* like\_count: Integer, number of likes.
* comment\_count: Integer, number of comments.
* published\_at: DateTime, original upload time.

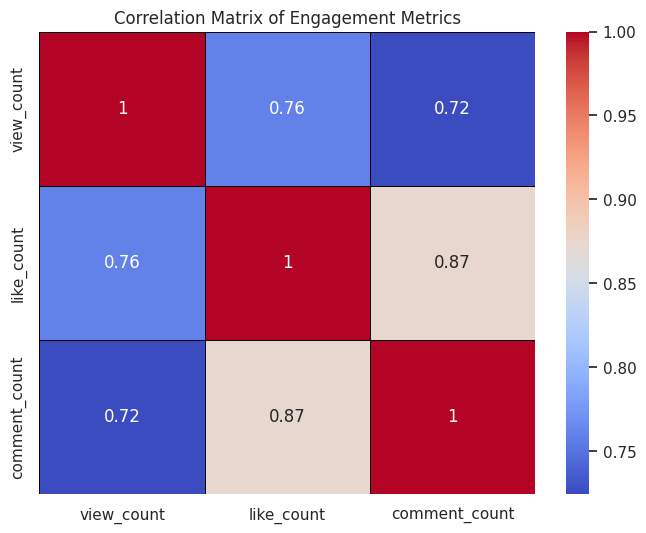
**Data Quality:**Some missing values were observed in the description column, and categorical inconsistencies were identified in the tags. These were addressed during the preprocessing stage.

***DATA CLEANING and PREPROCESSING***

Missing values in the description column were filled with "No description," and the published\_at column was converted to DateTime format for easier analysis. Additionally, tags were converted from string representations to lists to allow for more accurate handling of tag data.

# **EXPLORATORY DATA ANALYSIS (EDA)**

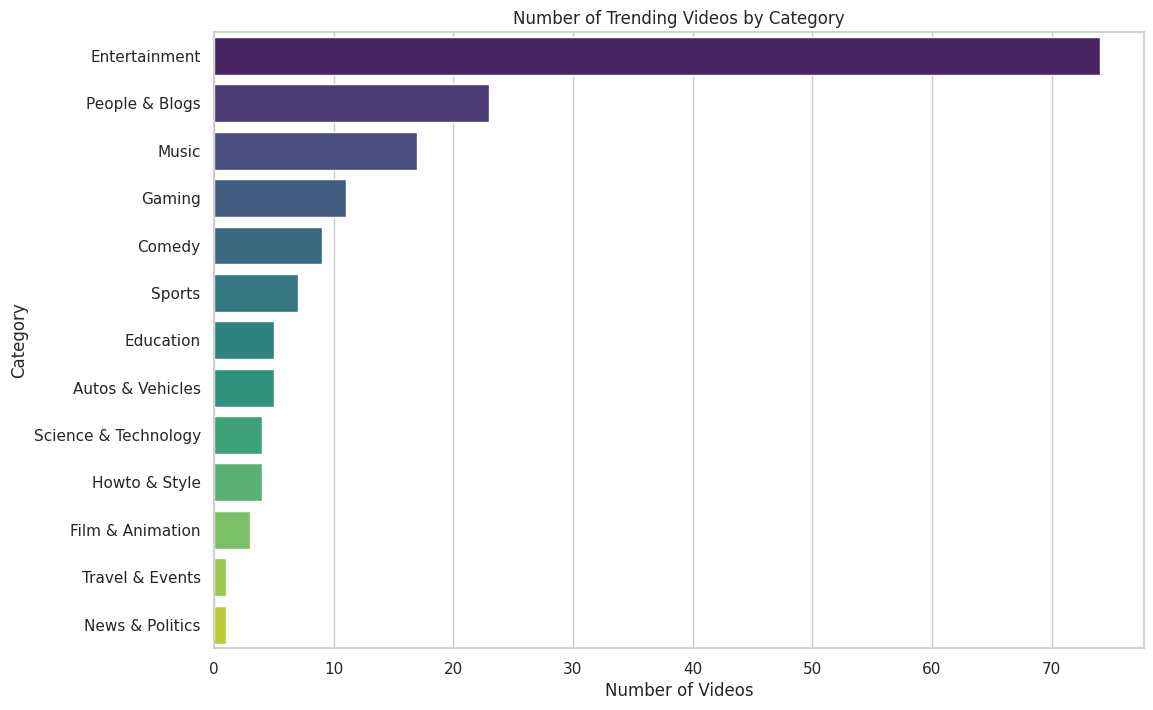
1. **Descriptive Statistics:** The dataset summary reveals that the average view count is approximately 2.1 million, with a standard deviation of around 4.36 million, indicating significant variation in viewership. Similarly, the mean like count is around 59,400, and the comment count has a mean of about 3,185, with both showing considerable spread, as seen in their respective standard deviations.
2. **Distribution Analysis:** The histograms and Kernel Density Estimate (KDE) plots for view counts, like counts, and comment counts indicate a highly skewed distribution, where most videos receive relatively low engagement. In contrast, a small number of videos achieve exceptionally high views, likes, and comments. This highlights that only a few trending videos dominate the engagement landscape, while the majority of videos experience limited interaction.
3. **Correlation Analysis:** The correlation matrix reveals strong positive relationships between the engagement metrics of view count, like count, and comment count. The highest correlation is observed between likes and comments (0.87), followed by views and likes (0.76), and views and comments (0.72). These correlations suggest that videos with more views are likely to garner more likes and comments, making any one of these metrics a potential predictor for others when evaluating video engagement.



# **RESULTS & DISCUSSION**

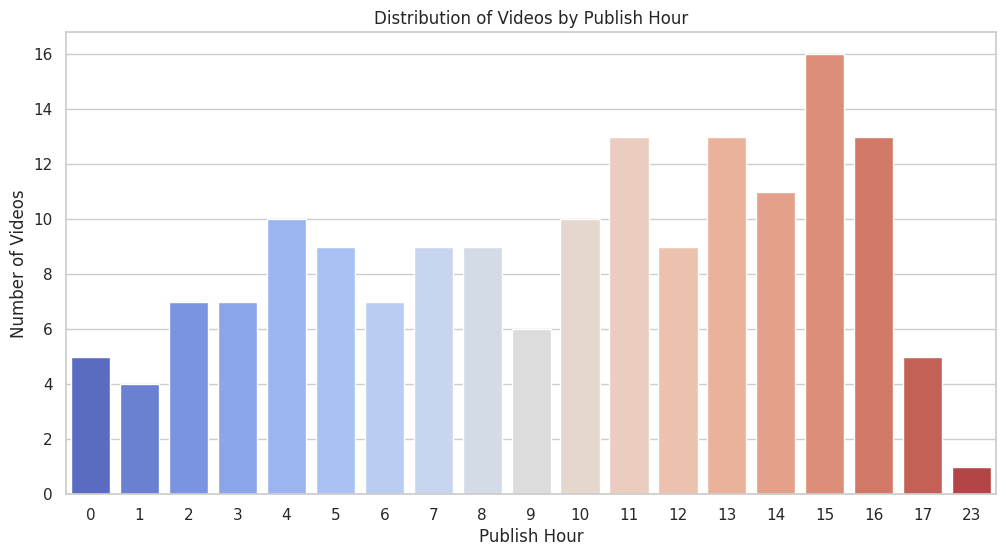
**Key Findings:**

* **Engagement Correlation:** Videos with higher view counts tend to receive more likes and comments, with the strongest relationship observed between likes and comments (0.87). This suggests that videos with high engagement across one metric often perform well across others.
* **Skewed Distribution:** The distribution of view counts, like counts, and comment counts are highly skewed, with most videos receiving lower engagement. Only a small subset of videos attracts significantly higher views, likes, and comments, indicating that a few videos dominate the engagement landscape.
* **Impact of Publish Hour:** While most videos(popularity in India Entertainment, People & Blogs, and Music categories have the highest number of trending videos) are published between 11 AM and 4 PM, which could be an optimal window for uploads, the published hour alone doesn't seem to strongly correlate with high view counts. A few videos achieve very high views at various times, suggesting other factors such as content quality and promotion play a more substantial role in video popularity.



**Implications:**

* The strong correlation between likes and comments indicates that driving more likes could help increase comments, suggesting that focusing on increasing one form of engagement may help boost overall interactions.
* The skewed distribution of engagement metrics implies that content creators should aim to produce videos that capture a larger audience, as a few videos generate the bulk of the engagement.
* While publishing videos during peak hours (11 AM - 4 PM) may be beneficial, creators should also consider other factors, such as content relevance and promotion, to maximize views.



# 

# **CONCLUSION & FUTURE WORK**

1. Encourage Engagement: Promoting likes and comments can significantly boost engagement metrics, as these interactions are strongly correlated with higher view counts.
2. Focus on Video Length: Creating longer videos (60 minutes) is linked with higher engagement, particularly in terms of views, likes, and comments.
3. Optimize Upload Timing: Uploading videos during peak hours (11 AM – 4 PM) can help maximize initial views and engagement, leveraging the high activity period of the platform.

# ***REFERENCES***

1. [Google Cloud Console](https://console.cloud.google.com/welcome?project=precise-data-392110&pli=1)
2. [YouTube Data Collection](https://ieeexplore.ieee.org/abstract/document/9150367)
3. [Content marketing strategy of branded YouTube channels](https://www.tandfonline.com/doi/abs/10.1080/16522354.2020.1783130)
4. [API Reference](https://developers.google.com/youtube/v3/docs)